Developmental Biology **242,** 188–203 (2002)
doi:10.1006/dbio.2002.0538, available online at http://www.idealibrary.com on IDEAL®

Gene Expression Profiles in Tadpole Larvae of *Ciona intestinalis*

Takehiro Kusakabe,*[,1] Reiko Yoshida,* Isao Kawakami,* Rie Kusakabe,* Yasuaki Mochizuki,† Lixy Yamada,† Tadasu Shin-i,‡ Yuji Kohara,‡ Nori Satoh,† Motoyuki Tsuda,* and Yutaka Satou†

**Department of Life Science, Himeji Institute of Technology, 3-2-1 Kouto, Hyogo 678-1297, Japan; †Department of Zoology, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan; and ‡Laboratory of Developmental Genetics, National Institute of Genetics, Mishima 411-8540, Japan*

A set of 12,779 expressed sequence tags (ESTs), both the 5′-most and 3′-most ends, derived from *Ciona intestinalis* tadpole larvae was categorized into 3521 independent clusters, from which 1013 clusters corresponding to 9424 clones were randomly selected to analyze genetic information and gene expression profiles. When compared with sequences in databases, 545 of the clusters showed significant matches (P < E-15) with reported proteins, while 153 showed matches with putative proteins for which there is not enough information to categorize their function, and 315 had no significant sequence similarities to known proteins. Sequence-similarity analyses of the 545 clusters in relation to the biological functions demonstrated that 407 of them have functions that many kinds of cells use, 104 are associated with cell–cell communication, and 34 are transcription factors or other gene-regulatory proteins. Sequence prevalence distribution analysis demonstrated that more than one-half of the mRNAs are rare mRNAs. All of the 1013 clusters were subjected to whole-mount *in situ* hybridization to analyze the gene expression profile in the tadpole larva. A total of 361 clusters showed expression specific to a certain tissue or organ: 96 showed epidermis-specific expression, 60 were specific to the nervous system, 108 to endoderm, 34 to mesenchyme, 5 to trunk lateral cells, 4 to trunk ventral cells, 23 to notochord, 28 to muscle, and 3 to siphon rudiments. In addition, 190 clusters showed expression in multiple tissues. Moreover, nervous system-specific genes showed intriguing expression patterns dependent on the cluster. The present study highlights a broad spectrum of genes that are used in the formation of one of the most primitive chordate body plans as well as for the function of various types of tissue and organ and also provides molecular markers for individual tissues and organs constituting the *Ciona* larva.

***Key Words: Ciona*; tadpole larvae; EST analysis; gene expression profiles; specific genes.**

INTRODUCTION

The ascidian tadpole larva consists of approximately 2600 cells. The tadpole is organized into a trunk and tail. The trunk contains a dorsal central nervous system (CNS) with two sensory organs (otolith and ocellus), endoderm, mesenchyme, trunk lateral cells (TLCs), and trunk ventral cells (TVCs). The tail contains a notochord flanked dorsally by the nerve cord, ventrally by the endodermal strand, and bilaterally by three rows of muscle cells. The entire surface of the larva is covered with an epidermis. This configuration of the ascidian tadpole represents one of the most simplified and primitive chordate body plans (Satoh, 1994, 1999, 2001; Satoh and Jeffery, 1995; Di Gregorio and Levine, 1998; Nishino and Satoh, 2001).

Ciona intestinalis, a species used by researchers throughout the world, has a small genome of about 1.6×10^8 bp/haploid containing approximately 15,500 genes (Simmen *et al.*, 1998). Large-scale cDNA analyses of gene expression profiles would facilitate investigation of the expression and function of developmental genes in embryos and larvae of this primitive chordate. In collaboration with the *C. intestinalis* cDNA project consortium members, our laboratories have conducted comprehensive studies of gene expression profiles in fertilized eggs (Nishikata *et al.*, 2001), 32- to 110-cell-stage embryos (Fujiwara *et al.*, 2002), tailbud embryos (Satou *et al.*, 2001), larvae (present study), and

[1] To whom correspondence should be addressed. Fax: +81-791-58-0197. E-mail: tgk@sci.himeji-tech.ac.jp.

0012-1606/02 $35.00

TABLE 1
The Number of Different Genes Expressed Per Class

	Class		No. of clusters	No. of clones	Clones/Cluster
	A. Functions that many kinds of cells use				
AI	Transport and binding proteins for ions and other small molecules		24	132	5.5
AII	RNA processing, polymerizing, splicing and binding proteins and enzymes		49	290	5.9
AIII	Cell replication, histones, cyclins and allied kinases, DNA polymerases, topoisomerases, DNA modification		22	106	4.8
AIV	Cytoskeleton and membrane proteins		65	1535	23.6
AV	Protein synthesis cofactors, tRNA synthetase, ribosomal proteins		88	1816	20.6
AVI	Intermediary metabolism and catabolism enzymes		98	646	6.6
AVII	Stress response, detoxification, and cell defense proteins		10	121	12.1
AVIII	Protein degradation and processing, proteases		28	91	3.3
AIX	Transportaion and binding proteins for proteins and other macromolecules		23	75	3.3
		Total	407	4812	11.8
	B. Cell–cell communication				
BI	Signaling receptors including cytokine and hormone receptors, and signaling ligands		10	29	2.9
BII	Intracellular signal transduction pathway molecules, including kinases and signal intermediates		68	273	4.0
BIII	Extracellular matrix proteins and cell adhesion		26	593	22.8
		Total	104	895	8.6
	C. Transcription factors and other gene regulatory proteins				
CI	Sequence-specific DNA-binding proteins		20	144	7.2
CII	Non-DNA-binding proteins that have positive or negative roles		6	22	3.7
CIII	Chromatin proteins other than AIII with regulatory function		8	45	5.6
		Total	34	211	6.2
	D. Not enough information to classify				
DI	Not enough information to classify		153	1625	10.6
DII	No significant similarities to known proteins		315	1881	6.0
		Total	468	3506	7.5
		Total	1013	9424	9.3

young adults (Ogasawara *et al.*, submitted) of *C. intestinalis.* As of the end of June 2001, about 22,000 expressed sequence tags (ESTs) (11,000 clones) had been determined for the tailbud embryos, and all of the 3′ ESTs are accessible in the DDBJ (GenBank/EMBL) database. The expression profiles of fully characterized genes at the tailbud stage are described at the web site http://ghost.zool.kyoto-u.ac.jp (Satou *et al.*, 2001, 2002). In addition, both the 3′- and 5′-end ESTs of about 16,000, 14,500, 12,000, and 22,000 clones have been determined for *Ciona* fertilized eggs, 32- to 110-cell-stage embryos, larvae, and young adults, respectively. The present study focused on the gene expression profiles of tadpole larvae, and here we show the results for 1013 independent clusters.

MATERIALS AND METHODS

Biological materials. *C. intestinalis* were cultivated at the Maizuru Fisheries Research Station of the Kyoto University and also collected near Murotsu, Hyogo, Japan. Eggs and sperm were obtained surgically from the gonoduct. After insemination, eggs were dechorionated by immersing them in seawater that contained 1.3% sodium thioglycolate (Wako Pure Chem. Ind. Ltd., Osaka, Japan) and 0.065% actinase E (Kaken Pharm. Co. Ltd., Tokyo, Japan). After the eggs were washed, they were reared in agar-coated dishes with Millipore-filtered seawater (MFSW) containing 50 μg/ml streptomycin sulfate at room temperature (18–20°C). Tadpole larvae hatched at about 17 h of development. After several hours of free-swimming period, specimens were collected for RNA isolation and whole-mount *in situ* hybridization.

cDNA library construction. Total RNA was isolated by the acid guanidinium thiocyanate–phenol–chloroform method (Chomczynski and Sacchi, 1987). Poly(A)$^+$ RNA was purified twice by using Oligotex beads (Roche Japan, Tokyo). Poly(A)$^+$ RNA was converted to double-stranded cDNA that contained an *Eco*RI site at the 5′-end and an *Xho*I site at the 3′-end using a cDNA synthesis kit (Stratagene, La Jolla, CA). The cDNAs were ligated to pBSII-SK(−) digested with *Eco*RI and *Xho*I and electroporated into XL-1 Blue MRF′ bacteria (Stratagene). The library was arrayed in 384-well plates in a Genetix Q-Pix robot.

TABLE 2
EST Sequence Similarities, Gene Description, and Probability of Occurrence by Chance

Category	Cluster ID	Accession No.	Database entry name	Organism	*P*
AI	00121	AF221690	voltage-dependent anion channel	*Squalus acanthias*	3E-59
	00205	AAA48982	Na,K-ATPase α 3-subunit	*Gallus gallus*	1E-76
	00452	CAA68743	calsequestrin homolog	*Gallus gallus*	6E-67
	00885	CAA73906	calmodulin	*Ciona intestinalis*	1E-80
	01095	AAA02625	(Na+,K+)-ATPase-β-2 subunit	*Gallus gallus*	4E-20
	01335	S47565	calcium-binding protein BDR-1	*Homo sapiens*	3E-88
	01661	BAA01643	H(+)-transporting ATPase	*Rattus norvegicus*	2E-25
	01945	Q15012	GOLGI 4-TRANSMEMBRANE SPANNING TRANSPORTER MTP	*Homo sapiens*	2E-20
	02049	AAA40991	calcium transporting ATPase	*Rattus norvegicus*	1E-88
	02159	AAB04104	ADP/ATP carrier protein	*Anopheles gambiae*	4E-18
	02538	P36609	NEURONAL CALCIUM SENSOR 2	*Caenorhabditis elegans*	3E-32
AII	00042	AAA36649	pre-mRNA splicing factor protein	*Homo sapiens*	8E-62
	00110	CAA46522	poly(A) binding protein	*Mus musculus*	2E-72
	00139	AAC72743	coding region determinant binding protein	*Mus musculus*	5E-46
	00158	BAA77512	cold-inducible RNA-binding protein	*Ciona intestinalis*	9E-24
	00366	AAD27610	nucleolar protein NOP5/NOP58	*Homo sapiens*	5E-39
	00406	AAA60373	Y-Box factor	*Aplysia californica*	3E-37
	00554	AAD38877	p68 RNA helicase	*Molgula oculata*	E-109
	00714	AAA49949	ribonucleoprotein	*Xenopus laevis*	4E-45
	00865	BAB03404	PEM-3	*Ciona savignyi*	7E-76
	01033	CAA58703	SP62_HUMAN; SAP 62; SF3A66	*Homo sapiens*	E-100
	01538	CAA45354	snRNP C	*Xenopus laevis*	5E-25
	01808	AAF37578	serine-arginine-rich splicing regulatory protein SRRP86	*Rattus norvegicus*	6E-31
	02342	AAF72188	snRNP-associated protein; SmB	*Danio rerio*	8E-32
AIII	00029	CAA61665	cyclin D2	*Xenopus laevis*	3E-55
	00109	CAA29061	histone H2 A.F/Z	*Strongylocentrotus purpuratus*	3E-53
	00211	S47348	histone H1.0	*Rattus norvegicus*	9E-17
	00342	AAD11940	mitotic checkpoint protein kinase BUB1B	*Mus musculus*	5E-35
	00483	CAA68458	histone H3.3B (AA 1–136)	*Gallus gallus*	1E-70
	01025	AAC00504	erythrocyte histone deacetylase	*Gallus gallus*	E-114
	01271	AAC69366	Cdc6-related protein	*Xenopus laevis*	3E-51
	01786	AAC34392	PAS1	*Takifugu rubripes*	1E-75
	02009	AAB09784	replication factor C, 36-kDa subunit	*Homo sapiens*	4E-81
	02052	BAA25400	CsCDC42	*Ciona savignyi*	E-106
AIV	00025	BAA23596	CsMA-I	*Ciona savignyi*	E-127
	00031	AAA59890	myosin regulatory light chain	*Homo sapiens*	2E-52
	00046	AAC28357	cytoskeletal actin 1	*Molgula occulta*	E-110
	00086	CAA26203	β-tubulin	*Homo sapiens*	3E-86
	00173	AAD09271	troponin 1	*Ciona intestinalis*	5E-92
	00189	AAA86910	fast-twitch myosin light chain 1	*Bos taurus*	1E-37
	00259	CAC03999	intermediate filament protein C	*Styela clava*	E-58
	00327	BAA08111	embryonic muscle myosin heavy chain	*Halocynthia roretzi*	3E-49
	00372	AAA48656	capping protein α 2 isoform	*Gallus gallus*	6E-16
	00608	CAA45469	tropomyosin	*Ciona intestinalis*	3E-48
	00982	CAA62350	α II spectrin	*Rattus norvegicus*	E-101
	00988	AAA60580	β-spectrin	*Homo sapiens*	2E-40
AV	00010	AAB26418	ribosomal protein 49	*Drosophila persimilis*	4E-40
	00045	CAA40592	ribosomal protein S1a	*Xenopus laevis*	5E-95
	00075	CAA63732	ribosomal protein L10a	*Rattus norvegicus*	8E-68
	00080	AAA50025	translational elongation factor-1 α	*Danio rerio*	6E-88
	00227	AAF64459	ribosomal protein L18	*Tilapia mossambica*	3E-71
	00514	PI7008	40S RIBOSOMAL PROTEIN S16	*Rattus norvegicus*	5E-55
	00693	A47151	ethionine adenosyltransferase (EC 2.5,1.6)	*Mus musculus*	2E-44
	01228	CAA73167	translation initiation factor eIF4A I	*Xenopus laevis*	3E-61
	01237	A53221	acidic ribosomal protein P1	*Polyorchis penicillatus*	8E-22
	01820	AAC38014	ribosomal protein S6	*Xenopus laevis*	6E-56

TABLE 2—*Continued*

Category	Cluster ID	Accession No.	Database entry name	Organism	*P*
AVI	00092	CAA39047	ornithine decarboxylase	*Homo sapiens*	1E-58
	00114	BAA05020	HrEpiB	*Halocynthia roretzi*	9E-77
	00138	BAA32086	natural killer cell enhancing factor	*Cyprinus carpio*	3E-78
	00140	AAA70333	adenylosuccinate synthetase	*Schizosaccharomyces pombe*	2E-44
	00155	AAA30359	S-adenosylmethionine decarboxylase	*Bos taurus*	2E-52
	00333	T02955	probable cytochrome P450 monooxygenase	*Zea mays*	9E-19
	00626	AAA30407	3-methyl-2-oxobutanoate dehydrogenase	*Bos taurus*	3E-64
	00661	AAD55350	golgi stacking protein homolog GRASP55	*Rattus norvegicus*	4E-81
	00881	AAD08929	glyceraldehyde-3-phosphate dehydrogenase	*Rattus norvegicus*	3E-89
	00927	AAD34720	glutamine synthetase	*Opsanus beta*	2E-58
	00957	BAA90476	thioredoxin peroxidase	*Ascaris suum*	8E-82
AVII	00331	CAA06233	heat shock cognate 70	*Gallus gallus*	2E-94
	00668	AAC48718	heat shock protein 90A	*Sus scrofa*	5E-90
	01307	AAA16364	amylo-1, 6-glucosidase/4-alpha-glucancotransferase	*Oryctolagus cuniculus*	2E-44
	02718	AAC96011	chaperonin containing t-complex polypeptide 1, η subunit; CCT η	*Homo sapiens*	7E-62
	02751	BAA90476	thioredoxin peroxidase	*Ascaris suum*	8E-82
	02793	AAD17774	arsenate resistance protein ARS2	*Homo sapiens*	1E-51
	04774	CAB94911	T-complex protein 1 δ subunit	*Gallus gallus*	2E-68
AVIII	00197	CAA80851	ubiquitin	*Phanerochaete chrysosporium*	E-114
	00509	AAC59636	carboxypeptidase H	*Lophius americanus*	8E-44
	00689	BAA11338	proteasome subunit p42 protein	*Homo sapiens*	2E-91
	00242	BAA03970	cathepsin L precursor	*Sarcophaga peregrina*	2E-68
	00264	CAA57512	XSUGI	*Xenopus laevis*	2E-67
	00267	CAA42568	ubiquitin extension protein	*Drosophila melanogaster*	1E-62
	00308	B42856	ubiquitin carrier protein E2	*Homo sapiens*	2E-62
	00346	AAD04181	β-transducin repeat containing protein	*Mus musculus*	6E-57
	00690	AAA82930	methionine aminopeptidase	*Homo sapiens*	2E-86
	01922	AAC69605	ubiquitin-conjugating enzyme UBC7	*Rattus norvegicus*	4E-65
	02209	BAA19748	26S proteasome subunit p44.5	*Homo sapiens*	1E-45
AIX	00022	BAA05019	HRSec61 protein	*Halocynthia roretzi*	8E-86
	00137	CAA04754	Rer1 protein	*Homo sapiens*	5E-62
	00149	CAA45217	TRAM-protein	*Canis familiaris*	4E-35
	00154	AAB18733	α-NAC, nonmuscle form	*Mus musculus*	8E-48
	00290	A42286	ERD-2-like protein, ELP-1	*Homo sapiens*	8E-96
	00389	CAA37662	glycoprotein 25L	*Canis familiaris*	3E-50
	00732	AAC79495	epsin 2	*Rattus norvegicus*	8E-71
	00798	CAA34386	SRP 54K subunit (AA 1–504)	*Mus musculus*	6E-82
	00854	AAC52154	transitional endoplasmic reticulum ATPase	*Rattus norvegicus*	6E-57
	01494	AAD13577	VAMP-associated protein B	*Homo sapiens*	1E-51
	03270	S11276	α-adaptin c	*Rattus norvegicus*	1E-29
BI	00520	CAA93803	hepatocyte growth factor (HGF) precursor protein	*Homo sapiens*	2E-38
	03264	AAC52272	complexin II	*Mus musculus*	5E-20
	03282	CAA62565	interleukin-8 receptor type B	*Macaca mulatta*	2E-20
	03293	AAC12877	secreted frizzled-related protein	*Homo sapiens*	1E-44
	04871	AAB66600	γ-aminobutyric-acid receptor rho-2B subunit	*Morone americana*	2E-23
	04933	AAB67639	connective tissue growth factor XCTGF	*Xenopus laevis*	8E-25
	06713	AAD32622	P opsin	*Anolis carolinensis*	1E-28
BII	00002	S39543	GTP-binding protein	*Mus musculus*	4E-56
	00060	AAB81617	receptor for activated protein kinase C	*Danio rerio*	1E-66
	00321	BAA92185	β-catenin protein	*Ciona intestinalis*	E-102
	00340	BAA92186	glycogen synthase kinase protein	*Ciona intestinalis*	6E-74
	00757	AAC35748	casein kinase 1 α isoform	*Gallus gallus*	E-108
	00777	AAD51932	RNA-binding protein isoform G3BP-2a	*Homo sapiens*	1E-52
	00931	CAB09533	RING3	*Homo sapiens*	1E-47
	00993	AAC48715	calcium/calmodulin-dependent protein kinase II δ 2-subunit	*Sus scrofa*	3E-72
	01063	AAC18960	GTPase cRac1A	*Gallus gallus*	1E-87

TABLE 2—*Continued*

Category	Cluster ID	Accession No.	Database entry name	Organism	*P*
	01545	AAB64904	Snf1p: serine/threonine protein kinase; CAI: 0.19	*Saccharomyces cerevisiae*	1E-20
	01990	P30676	GUANINE NUCLEOTIDE-BINDING PROTEIN G(I), α SUBUNIT	*Asterina pectinifera*	3E-91
	02082	AAD02277	M-Ras	*Mus musculus*	7E-61
BIII	00035	CAB38568	fibulin-5	*Homo sapiens*	8E-23
	00093	BAA94972	type I collagen α	*Xenopus laevis*	1E-31
	00113	AAB92586	laminin β 2-like chain	*Gallus gallus*	7E-22
	00289	CAB96130	claudin	*Halocynthia roretzi*	2E-24
	00299	AAA48707	α 1 type XI collagen	*Gallus gallus*	2E-45
	00561	BAA03127	entactin/nidogen	*Halocynthia roretzi*	2E-58
	00837	AAF44681	collagen IV α1 chain	*Gallus gallus*	7E-90
	01475	P27393	COLLAGEN α 2(IV) CHAIN PRECURSOR	*Ascaris suum*	2E-61
	01732	AAA02891	tight junction (zonula occludens) proteins ZO-1	*Homo sapiens*	9E-52
	02194	JC5951	integrin α 7 chain variant	*Homo sapiens*	3E-38
CI	00220	AAC60126	xGCNF	*Xenopus laevis*	7E-19
	00250	BAA08722	As-MEF2	*Halocynthia roretzi*	8E-66
	00255	A53184	myc far upstream element-binding protein	*Homo sapiens*	1E-19
	00319	AAF60348	Cdx	*Herdmania curvata*	8E-17
	00350	AAD24209	CCCH zinc finger protein C3H-3	*Xenopus laevis*	9E-34
	00517	AAD00562	transcriptional regulator Sox-11B	*Danio rerio*	1E-20
	00905	BAA96136	PBX1B	*Gallus gallus*	5E-39
	01659	CAA69928	nuclear orphan receptor ROR-β	*Gallus gallus*	3E-49
	02043	AAA20993	NF45 protein	*Homo sapiens*	9E-77
	02215	AAB03529	zinc finger protein	*Mus musculus*	1E-21
	02388	BAA31223	BAZF	*Mus musculus*	4E-18
	02710	BAA34234	hypoxia-inducible factor-1 α	*Gallus gallus*	6E-25
	02711	AAD34170	four and a half LIM-domain protein DRAL	*Mus musculus*	2E-50
	03308	A32608	thyroid hormone receptor-related protein Rev-ErbA α	*Homo sapiens*	1E-33
	04708	AAF05623	orphan nuclear receptor TECδC short isoform	*Mus musculus*	4E-34
	04797	AAF51966	cas gene product	*Drosophila melanogaster*	4E-16
	05534	AAF34700	cysteine-rich protein NFX-1	*Mus musculus*	7E-60
	06639	CAC10403	iroquois homeobox protein 6	*Mus musculus*	2E-41
CII	00684	AAA61194	transducin-like enhancer protein	*Homo sapiens*	9E-38
	02949	BAA85180	C-terminal binding protein 1	*Mus musculus*	1E-89
	04071	P25655	GENERAL NEGATIVE REGULATOR OF TRANSCRIPTION SUBUNIT 1	*Saccharomyces cerevisiae*	3E-23
CIII	00488	Q08945	STRUCTURE-SPECIFIC RECOGNITION PROTEIN 1 (SSRP1)	*Homo sapiens*	3E-61
	00531	CAA61938	ssrp2	*Drosophila melanogaster*	6E-54
	03260	AAC50693	SWI/SNF complex 155 KDa subunit	*Homo sapiens*	2E-38
	03266	P50238	CYSTEINE-RICH PROTEIN 1 (CYSTEINE-RICH INTESTINAL PROTEIN) (CRIP)	*Homo sapiens*	6E-29
	03366	AAB18950	Supt6h	*Mus musculus*	3E-50
	03589	AAC51735	ALR	*Homo sapiens*	8E-49
	07645	BAA89208	bromodomain PHD finger transcription factor	*Homo sapiens*	9E-45

EST sequencing, and clustering and similarity search of ESTs. Clones were picked up from the 384-well plates and cDNA inserts were amplified by using PCR. Successful amplifications were confirmed by using agarose gel electrophoresis. After the PCR products were purified, their sequences were determined by conventional procedures using the big-dye terminators on ABI 3700 autosequencers at the Academia DNA Sequencing Center (National Institute of Genetics, Mishima, Japan). The primer for 3′-sequencing was the anchored oligo(dT) primers (5′-$(dT)_{17}$-V-3′) and that for 5′-sequencing was BS740 (5′-CCGCTCTAGAACTAGTG-3′).

Using 3′-most sequence tags, clones were grouped into clusters, each of which contained cDNA clones corresponding to the same gene. The DDBJ DNA database and protein database (DAD) were searched with the 5′-most sequence tags using the BLAST algorithm (BLASTN and BLASTX). We categorized clusters into several groups according to their functions that were predicted from the BLASTX results (see Results and Discussion). The classification done by Lee *et al.* (1999) was referred to extensively.

Whole-mount in situ hybridization. The gene expression profiles of tadpole larvae were examined by whole-mount *in situ* hybridization. Specimens were fixed in 4% paraformaldehyde in

0.1 M MOPS (pH 7.5), 0.5 M NaCl at 4°C for 16 h, prior to storage in 80% ethanol at −30°C. For convenience, we applied a method of whole-mount *in situ* hybridization using DNA probes to ascidian larvae. DNA probes were synthesized basically as described in Tabara *et al.* (1996) with several modifications (Satou *et al.*, 2001). First, cDNA inserts were PCR-amplified with SK and T7 primers. Each 10-μl reaction mixture contained 1.4 ng of template DNA, 0.1 μM primers, 0.2 mM of the four dNTPs, 1.5 mM $MgCl_2$, 1× buffer, and 0.25 U *Taq* DNA polymerase (TOYOBO). Reactions proceeded through 35 cycles (30 s at 94°C, 30 s at 55°C, 3 min at 72°C). DIG-labeled DNA probes were also synthesized by PCRs. Each 10-μl reaction mixture contained 1 μl of first PCR product, 0.0525 mM DIG–dUTP (Roche Japan), 0.0975 mM dTTP, 0.15 mM dATP, 0.15 mM dCTP, 0.15 mM dGTP, 1.5 mM $MgCl_2$, 0.53 μM $(dT)_{17}dG$, 0.53 μM $(dT)_{17}dC$, 2.13 μM $(dT)_{17}dA$, 1× buffer, and 0.5 U *Taq* DNA polymerase (TOYOBO). Reactions proceeded through 30 cycles (30 s at 95°C, 1 min at 45°C, 1 min at 72°C). After synthesis, free nucleotides and primers that were not incorporated were removed by ethanol precipitation.

Washes of specimens, proteinase K treatment, postfixation, prehybridization, hybridization, and color reaction of specimens were carried out according to the method described in Satou *et al.* (2001). Some of the specimens were dehydrated in a graded series of ethanol, then cleared in a 1:2 mixture of benzyl alcohol:benzyl benzoate.

TABLE 3

Overall View of Specific Expression Patterns of Genes in *C. intestinalis* Larva

Tissue	Genes specifically expressed in a single tissue No. of clusters	Genes specifically expressed in multiple tissues No. of clusters
Epidermis	96	157
Nervous system	60	185
epidermal neuron	3[a]	12[a]
brain	18[a]	128[a]
papilla	16[a]	81[a]
ganglion	4[a]	71[a]
nerve cord	7[a]	75[a]
Endoderm	108	175
Endodermal strand	0	12
Mesenchyme	34	69
Trunk lateral cells	5	17
Trunk ventral cells	4	12
Atrial siphon	2	8
Oral siphon	1	8
Notochord	23	47
Muscle	28	50
Weak and ubiquitous expression	321	—
Not detected	141	—
(Multiple tissues)	190	—
Total	1013	—

[a] The number of genes which are expressed specifically in these cell types (the number is included in that of the nervous system).

RESULTS AND DISCUSSION

Overall Distribution of Sequences

In the present study, we determined the sequences of both the 5′-most and the 3′-most ends of a total of 12,779 cDNA clones derived from *C. intestinalis* tadpole larvae. The average length of inserts of cDNAs was 1.7 kb, ranging from 0.5 to 5.0 kb. We determined 400- to 500-nt sequences of both ends. The sequences of the 3′-most ends were compared with one another by using the FASTA program (Pearson and Lipman, 1988) to examine the overlap of clones. The threshold values for clustering were 150 in similarity score and 89% in sequence identity. This analysis categorized the 12,779 clones into 3521 independent clusters. We randomly selected 1013 clusters corresponding to 9424 clones for further analysis.

The sequences of the 5′-most ends of the 1013 clusters were subjected to BLASTX analysis, and this analysis showed that 545 clusters of these cDNAs were strong matches ($P <$ E-15) to previously identified proteins with distinct functions (Table 1). The frequency of matches was therefore 545/1013, or about 0.54. As summarized in Table 1, these 545 proteins were categorized into three major classes basically according to the classification scheme of Lee *et al.* (1999): Class A, which contained 407 proteins associated with functions that many kinds of cell use; Class B, which contained 104 proteins associated with cell–cell communication; and Class C, which contained 34 proteins that function as transcription factors or other gene-regulatory proteins. Furthermore, the 407 proteins of Class A were categorized into nine subclasses (AI–AIX): 24 proteins that function as transport and binding proteins for ions and other small molecules (AI); 49 proteins that function as RNA-processing, polymerization, splicing, and binding proteins and enzymes (AII); 22 proteins that function in cell replication (AIII); 65 proteins that are cytoskeleton and membrane proteins (AIV); 88 proteins that are protein synthesis cofactors, ribosomal proteins, or function in tRNA synthesis (AV); 98 proteins that function as enzymes for intermediary synthesis and catabolism (AVI); 10 proteins that are associated with stress response, detoxification, and cell defense (AVII); 28 proteins that function in protein degradation and processing or are proteases (AVIII); and 23 proteins that function in the transportation and binding of proteins or other macromolecules (AIX) (Table 1).

The 104 proteins of Class B were further classified into three subclasses (BI–BIII): 10 proteins that function as signal receptors and ligands (BI), 68 proteins that function as intracellular signal transduction pathway molecules (BII), and 26 proteins that are extracellular matrix proteins or are otherwise related to cell adhesion (BIII) (Table 1). The 34 proteins of Class C were further classified into three subclasses (CI–CIII): 20 proteins that function as sequence-specific DNA-binding proteins (CI), 6 proteins that function as non-DNA-binding proteins with positive or negative regulatory roles in gene regulation (CII), and 8 proteins that

TABLE 4
Relationship between Number of Specifically Expressed Genes and Their Classes

	Epidermis		Nervous system		Endoderm		Mesenchyme		Notochord		Muscle	
Class	No. of clusters	No. of clones	No. of clusters	No. of clones	No. of clusters	No. of clones	No. of clusters	No. of clones	No. of clusters	No. of clones	No. of clusters	No. of clones
AI	1	3	4	18	1	9	2	7	0	0	3	48
AII	9	31	0	0	6	66	2	48	0	0	1	3
AIII	1	4	0	0	7	36	0	0	1	2	0	0
AIV	4	462	3	14	2	4	1	2	1	2	13	721
AV	6	83	2	9	42	1314	6	73	2	5	0	0
AVI	2	11	4	16	8	65	4	20	4	16	2	47
AVII	0	0	0	0	4	64	0	0	0	0	0	0
AVIII	2	6	4	17	1	5	0	0	0	0	0	0
AIX	0	0	2	5	1	7	5	19	0	0	0	0
BI	1	5	2	8	0	0	1	3	0	0	0	0
BII	5	17	7	40	3	15	0	0	1	3	2	10
BIII	7	408	0	0	3	12	2	8	1	7	0	0
CI	2	10	1	1	3	13	1	3	0	0	0	0
CII	1	4	1	4	0	0	0	0	0	0	0	0
CIII	1	2	0	0	1	29	0	0	0	0	0	0
DI	18	863	9	38	10	84	3	14	3	12	1	3
DII	36	254	21	87	16	174	7	133	10	86	6	58

function as chromatin proteins (other than subclass AIII proteins) with regulatory functions (CIII) (Table 1).

Besides the 545 clusters with strong matches, 153 clusters were classified into subclass DI, with sequences that were matches to ESTs (mostly from *Caenorhabditis elegans*, *Mus musculus*, and *Homo sapiens*) or reported proteins for which not enough information is available to classify their functions (Table 1). The remaining 315 clusters were categorized into subclass DII, with no significant sequence similarities to known proteins (Table 1). We are now sequencing the entire nucleotides of these clusters to determine whether the clusters represent truly unknown genes or not.

Sequence Analysis of Selected cDNA Clones

Table 2 shows some examples of sequence analysis to identify cDNAs that encode especially strong candidates for proteins with a defined function. For example, clones in subclass BI (which consists of clones for signaling receptors, such as cytokine and hormone receptors and signal ligands) included cDNAs for hepatocyte growth factor precursor protein (cluster ID 00520; $P <$ 2E-38), secreted frizzled-related protein (03293; $P <$ 1E-44), connective tissue growth factor XCTGF (04933; $P <$ 8E-25), and P opsin (06713; $P <$ 1E-28). Subclass BII, consisting of clones for intracellular signal transduction pathway molecules such as kinases and signal intermediates, contained cDNAs for GTP-binding protein (cluster ID 00002; $P <$ 4E-56), receptor for activated protein kinase C (00060; $P <$ 1E-66), β-catenin protein (00321; $P <$ E-102), glycogen synthase kinase protein (00340; $P <$ 6E-74), and RING3 (00931; $P <$ 1E-47). Subclass BIII for extracellular matrix proteins or proteins otherwise related to cell adhesion contained cDNAs for fibulin-5 (cluster ID 00035; $P <$ 8E-23), claudin (00289; $P <$ 2E-24), entactin/nidogen (00561; $P <$ 2E-58), collagen IV α1 chain (00837; $P <$ 7E-90), and tight junction (zonula occludens) protein ZO-1 (01732; $P <$ 9E-52).

In addition, as members of the subclass CI of clones for sequence-specific DNA-binding proteins, cDNAs for Cdx (cluster ID 00319; $P <$ 8E-17), transcriptional regulator Sox-11B (00517; $P <$ 1E-20), NF45 protein (02043; $P <$ 9E-77), and iroquois homeobox protein (06639; $P <$ 2E-41) were identified. The cDNAs for transducin-like enhancer protein (00684; $P <$ 9E-38) and C-terminal-binding protein 1 (02949; $P <$ 1E-89) were identified as members of subclass CII for clones encoding non-DNA-binding proteins that perform positive and negative roles in gene regulation, and cDNAs for structure-specific recognition protein 1 (SSRP1) (cluster ID, 00488; $P <$ 3E-61), ssrp2 (00531; $P <$ 6E-54), and cysteine-rich protein 1 (cysteine-rich intestinal protein: 03266; $P <$ 6E-29) were identified as members of subclass CIII for clones encoding chromatin proteins.

Prevalence Distribution of mRNAs

Our library was not amplified, so the number of clones per cluster may represent the frequency of the respective mRNA in the *Ciona* larva. Thus, analysis of the number of clones constituting each cluster allowed us to determine the prevalence distribution of mRNAs expressed in the larva. The number of clones belonging to each subclass is

also shown in Table 1. The great majority of clusters were identified several times, and the average number of clones/cluster was 9.3 (9424/1013; Table 1). For example, it is evident that the numbers of clones/cluster of the subclasses AIV (cytoskeleton and membrane proteins) (1535/65 = 23.6), AV (protein synthesis cofactors, ribosomal proteins or proteins that function in tRNA synthesis) (1816/88 = 20.6), and BIII (extracellular matrix proteins or cell adhesion proteins) (593/26 = 22.8) were rather high compared with those of other subclasses (Table 1).

Detailed analysis of the prevalence distribution of mRNAs demonstrated that 162 of the 1013 clusters were identified only once, while 218 clusters were identified only twice, 168 clusters three times, and 101 clusters four times (data not shown). Therefore, it is likely that more than one-half of the mRNAs expressed in the *Ciona* larva are rare mRNAs.

In contrast, a gene (cluster ID 00046) which encodes a cytoplasmic actin was identified 418 times. In addition, cluster 00025 for a muscle actin, cluster 00419 for a protein of the subclass DI, and cluster 00065 for a protein of the subclass DI contained 302, 264, and 252 clones, respectively. Furthermore, cluster 00087 for a protein of subclass DI contained 230 clones, cluster 00085 for acid ribosomal protein P0 189 clones, cluster 00081 for 40S ribosomal protein 171 clones, cluster 00080 for EF-1α 167 clones, and cluster 00359 for troponin T 144 clones. The present data about the prevalence distribution of mRNAs suggest that more than one-half of the genes expressed in the larva correspond to rare messages and that a few genes, especially genes for muscle proteins, are very actively transcribed.

Spatial Expression Profiles of Genes

Overview. All of the 1013 clusters were subjected to analysis by whole-mount *in situ* hybridization to explore the overall gene expression profiles in *Ciona* larvae. As summarized in Table 3, 141 clusters showed no significant hybridization signals. This lack of gene expression patterns is presumably due to the presence of low levels of mRNAs in the larva, because most such clusters show a small number of clones/cluster. In addition, 321 clusters showed very weak and ubiquitous signals.

However, 361 of the 1013 clusters showed spatial expression patterns that were specific to a single tissue or organ (Table 3). Within these clusters, 96 genes were expressed specifically in epidermal cells, 60 genes were specific to the nervous system, including neuronal cells in the papilla (palps: an adhesive organ located at the anterior most part of the embryo) and epithelial sensory cells. Endoderm-specific expression was observed for 108 genes, the number of mesenchyme-specific genes was 34, that of TLC-specific genes was 5, and that of TVC-specific genes was 4. Notochord-specific expression was detected for 23 genes, and 28 genes were expressed specifically in muscle cells. In addition, we found three genes that were expressed in the siphon rudiments. Several examples of specific gene expression for each tissue will be described in the following section.

Table 4 and Fig. 1 show the relationships between the genes with single-tissue-specific expression and classes obtained from sequence data. This analysis highlights the characteristic gene expression profile for each type of tissue. Nearly one-half of the epidermis-specific genes with distinct functions are in Class A (Fig. 1A). However, Fig. 1A also shows that genes belonging to subclasses BII and BIII are extensively expressed in the epidermis. Of the genes that are specifically expressed in the nervous system, more than 60% are categorized into Class A (Fig. 1B). It is also evident that there is intensive expression of subclass BII genes for signal transduction molecules in the nervous system (Fig. 1B).

Figure 1C shows that, as was expected, many endoderm-specific genes are associated with protein synthesis (subclass AV). Many mesenchyme-specific genes are associated with RNA processing (AII), protein synthesis (AV), intermediary metabolism (AVI), and transport of proteins (AIX): more than 70% of mesenchyme-specific mRNAs are for these functions (Fig. 1D). Figure 1E shows that the most abundant mRNAs of genes that are expressed specifically in the notochord are associated with metabolism (subclass AVI), suggesting that ascidian notochord cells are metabolically active. Genes for subclasses BII and BIII are also expressed in the notochord (Fig. 1E). As was expected, many muscle-specific genes are associated with cytoskeleton proteins (subclass AIV), and this category of gene accounted for nearly 90% of the clones expressed in muscle (Fig. 1F).

Specific expression patterns. The following are descriptions of representative genes that show tissue-specific expression patterns. The spatial expression pattern for every gene appears in our web site (http://ghost.zool.kyoto-u.ac.jp).

EPIDERMIS. A single layer of 800 epidermal cells constitutes the outermost layer of the ascidian larva. Expression patterns of four representative epidermis-specific genes are shown in Figs. 2A–2D. Many epidermis-specific genes were expressed over the entire surface of the larva, as represented by cluster 01230, which shows no significant sequence similarity to known genes (Fig. 2A). However, several genes were differentially expressed in the trunk and tail epidermis. For example, cluster 00341, encoding a protein similar to calumenin, is expressed only in the tail epidermis (Fig. 2B). In contrast, cluster 00861, with no significant similarity to known genes, is only expressed in the trunk epidermis (Fig. 2C). Interestingly, cluster 00875, showing a significant similarity to *Homo sapiens* brain-specific angiogenesis inhibitor is expressed in both the trunk and tail epidermis, but the expression in the tail is the strongest at the base and weaker in the posterior part of the tail (Fig. 2D).

In *Halocynthia roretzi* embryos, cDNA clones for eight different epidermis-specific genes have been isolated and characterized (Ueki *et al.*, 1991; Ueki and Satoh, 1995; Ishida *et al.*, 1996), while *Cs-Epi1* and *Cs-Epi2* are two genes that are specifically expressed in *C. savignyi* embryonic

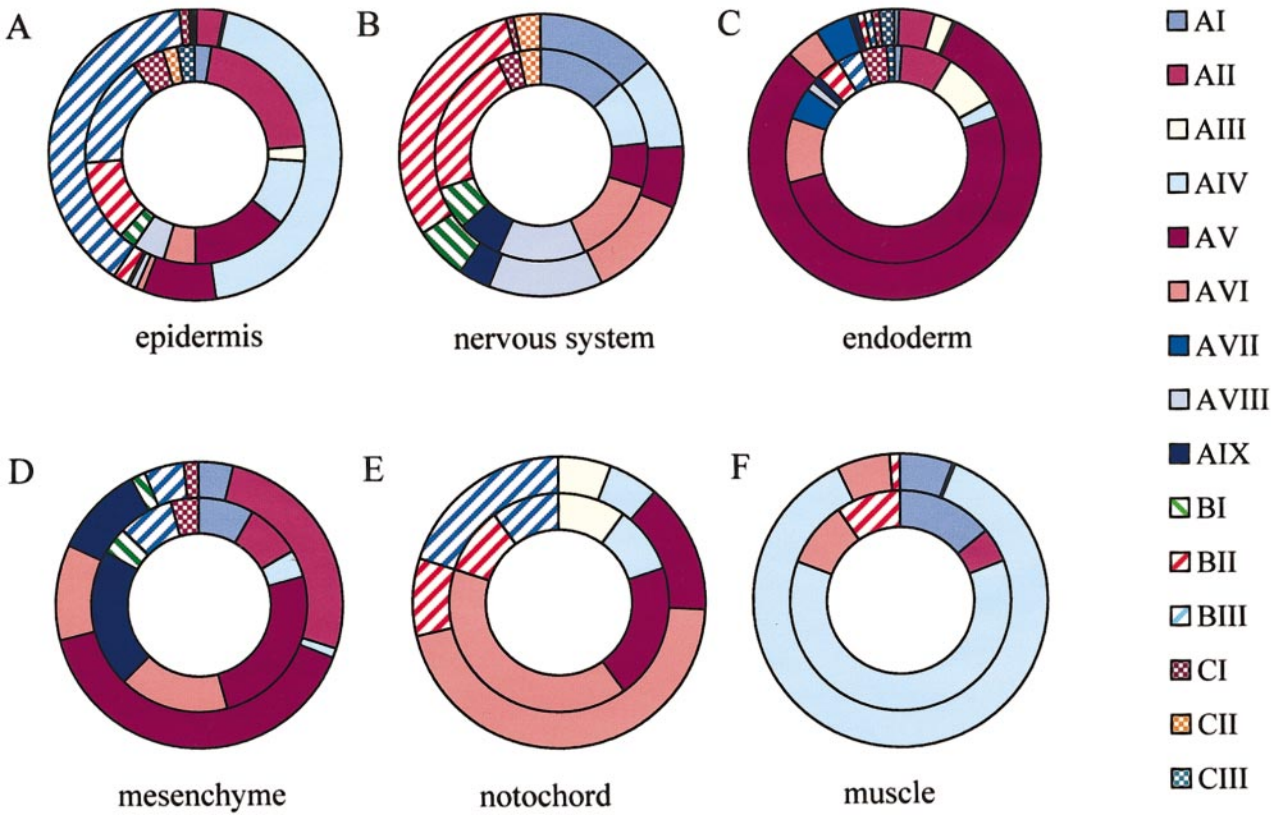

FIG. 1. Proportion of clones (outer circle) and clusters (inner circle) for the different classes of genes that are specifically expressed in (A) epidermis, (B) nervous system, (C) endoderm, (D) mesenchyme, (E) notochord, or (F) muscle. The classification of genes is shown in Table 1. Real numbers of genes are listed in Table 4.

epidermal cells (Chiba *et al.*, 1998). However, gene expression specific to trunk epidermis has not been observed in *Halocynthia* embryos (Ishida *et al.*, 1996) or in gene expression profiles of *C. intestinalis* tailbud embryos (Satou *et al.*, 2001). Genes displaying the trunk-specific expression may be involved in development of the adult epidermis or formation of the adult tunic, because the tail epidermis degenerates during metamorphosis of larvae to the adult form. Future studies on the expression and function of these trunk epidermis-specific genes will give insight into the molecular mechanisms of the development of the adult epidermis and tunic.

Ishida *et al.* (1996) reported that the spatiotemporal expression patterns of eight epidermis-specific genes of *Halocynthia* embryos could be categorized into four groups: genes expressed over the entire surface of the embryo, genes expressed only in the tail epidermis, genes expressed in the midline, and genes expressed in all epidermal cells except for midline cells. In the present analysis, the latter two types of expression pattern were not observed. This is in marked contrast to our previous analysis of gene expression profiles in *Ciona* tailbud embryos, in which all four patterns were observed (Satou *et al.*, 2001). The lack of expression patterns associated with the midline epidermis in the larva is consistent with the idea that epidermal cells along the midline of the tailbud embryo are specialized to form the larval fin (Ishida *et al.*, 1996), because the larval fin formation would be completed by the time of hatching.

NERVOUS SYSTEM. Formation of the nervous system of *Ciona* larvae has been studied in detail (Katz, 1983; Nicol and Meinertzhagen, 1991; Takamura, 1998). The nervous system comprises the CNS and peripheral sensory nervous system, consisting of approximately 350 cells. The anteriormost part of the larva is an adhesive organ or papilla (palps) formed by the differentiation of several neuronal cells. The mid-dorsal trunk region contains a brain vesicle with two sensory organs called the otolith and ocellus. The CNS extends via visceral ganglion in the neck region into the tail nerve cord, which consists of glial ependymal cells. Peripheral epithelial neuronal cells are distributed on the dorsal surface of the trunk as well as on the dorsal and ventral midlines of the tail epidermis.

Until recently, only two monoclonal antibodies have been used to monitor neuronal differentiation in *Ciona* embryos (Crowther and Whittaker, 1992; Takamura, 1998).

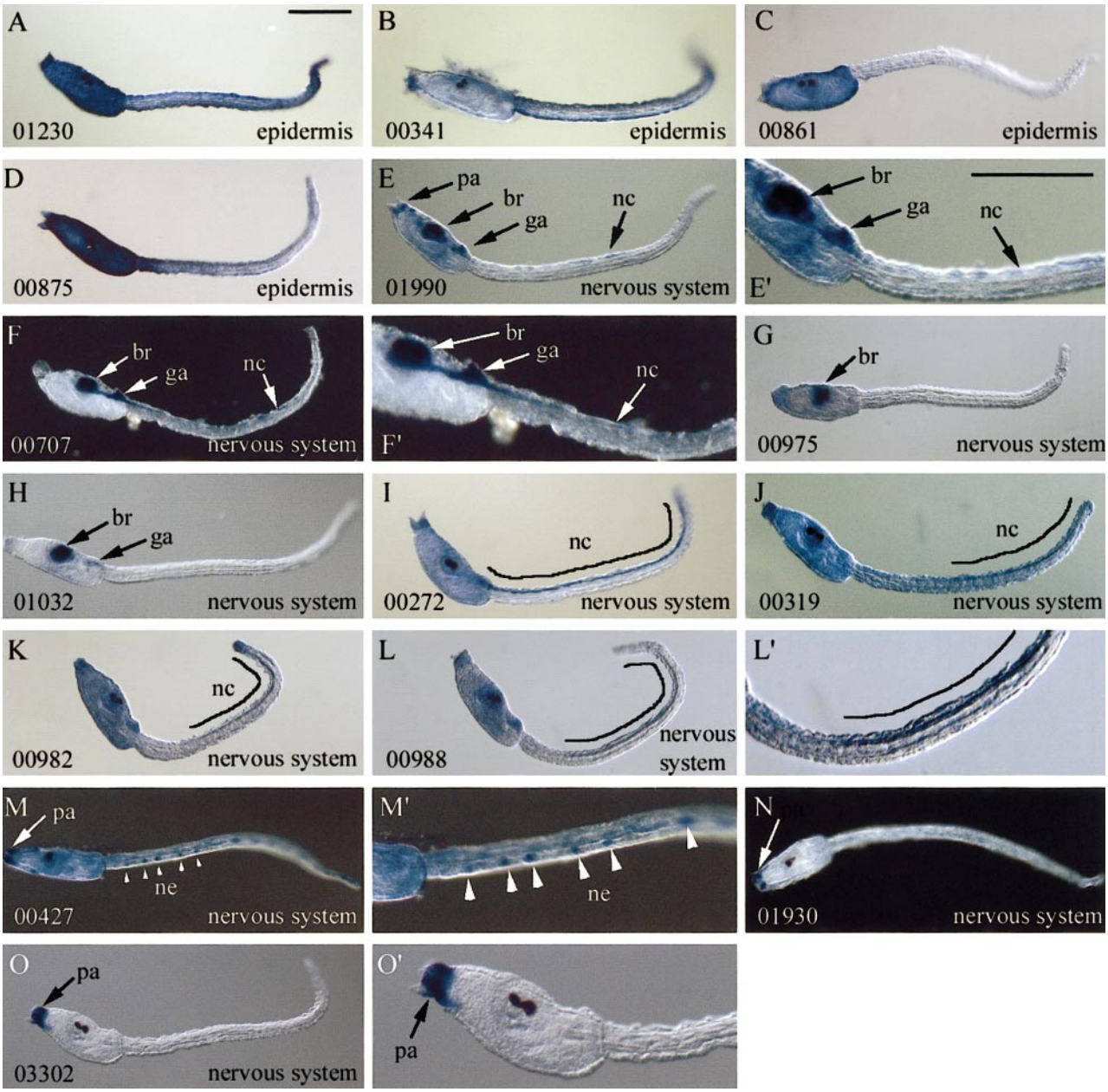

FIG. 2. Whole-mount *in situ* hybridization showing the expression profiles of genes specific to epidermis or nervous system in *C. intestinalis* larvae. The organ where the gene is specifically expressed is shown in the right bottom corner, and the clone names in the left bottom corner. (E′), (F′), (L′), (M′), and (O′) are enlargements of (E), (F), (L), (M), and (O), respectively. br, brain; ga, visceral ganglion; nc, nerve cord; pa, papilla. Black lines in (I–L) and (L′) indicate regions with gene expression within the nerve cord. Arrowheads in (M) and (M′) indicate epithelial sensory cells. Scale bars in (A) and (E′) are 100 μm and are applicable to all photographs. See text for details.

In the previous analysis of gene expression profiles of *C. intestinalis* tailbud embryos, 34 of 1213 genes examined were shown to be expressed exclusively in the nervous system (Satou *et al.*, 2001). Interestingly, we found a larger number of nervous system-specific genes in the present analysis: 60 of 1013 clusters were expressed exclusively in the nervous system of the larva (Table 3). For example, cluster 01990, encoding a G protein α subunit, is expressed in the CNS and the adhesive organ (Figs. 2E and 2E′). Both clusters 00707 (no known sequence similarity; Figs. 2F and 2F′) and 01098 (encoding synaptotagmin; data not shown) are also conspicuously expressed in the CNS. Cluster

00975, with no significant matches in the databases, is expressed only in the brain vesicle (Fig. 2G). Another example of brain vesicle-specific genes is cluster 01032, which encodes a protein similar to β-carotene, 15,15′-dioxygenase. This gene was expressed strongly in the brain vesicle and weakly in the visceral ganglion containing motor neurons (Fig. 2H).

The present study identified seven genes that are expressed exclusively in the nerve cord. For example, cluster 00272, showing no significant matches in the databases, was only expressed in the nerve cord (Fig. 2I). Cluster 00319 encodes a homolog of Caudal homeodomain protein. The *caudal* gene of *Halocynthia* embryos is expressed in posterior neuronal cells and posterior epidermal cells as well (Katsuyama *et al.,* 1999). The *caudal* gene of *Ciona* is only expressed in the posterior half of the nerve cord with an anteroposterior gradient of expression (Fig. 2J). Interestingly, two genes, 00982 and 00988 encoding α and β chains of spectrin, respectively, showed an expression pattern very similar to that of the *caudal* homolog (Figs. 2K, 2L, and 2L′). These genes may be downstream targets of the Caudal homeodomain protein.

In addition to CNS-specific genes, the present analysis identified genes expressed exclusively in the peripheral nervous system, which consists of epidermal neurons and neurons of the adhesive organ. For example, clone 00427, showing no significant similarity to known genes, is expressed in both types of peripheral neurons, but not in the CNS (Figs. 2M and 2M′). Clone 01930, encoding a protein similar to peptidyl prolyl *cis*-trans isomerase, is exclusively expressed in neurons of the adhesive organ (Fig. 2N). Interestingly, clone 03302, with no significant known sequence similarity, is exclusively expressed in the epidermis of the basal part of the adhesive organ (Figs. 2O and 2O′).

In contrast to organs such as muscle, epidermis, and endoderm, in which cell fates are autonomously determined by maternally inherited cytoplasmic factors, successive cell–cell interactions play central roles in cell-fate specification of the nervous system in ascidian embryos (Reverberi *et al.,* 1960; Okado and Takahashi, 1988; Nishida, 1991; Hudson and Lemaire, 2001). Therefore, specification and differentiation in the nervous system seem to occur relatively late during embryogenesis. Consequently, we have detected more nervous system-specific genes in larvae than in tailbud embryos (cf., Satou *et al.,* 2001). The nervous system of *Ciona* provides researchers with a simple and primitive model to analyze the developmental and functional complexities of the vertebrate CNS. Genes that were identified in the present study should be very useful for such studies in the future.

ENDODERM. The endodermal tissue of the *Ciona* tadpole larva constitutes the trunk endoderm and tail endodermal strand. The endoderm consists of about 500 cells. Because the tadpole of most ascidian species is the dispersal phase of their life cycle, tadpoles do not open their mouths prior to attachment to the substrate, which is an early event of metamorphosis. The histochemical detection of alkaline phosphatase (AP) activity is a conventional way to study endoderm differentiation (e.g., Whittaker, 1990), and a cDNA clone for the AP gene was isolated from *H. roretzi* (Kumano and Nishida, 1998) and *C. intestinalis* and *C. savignyi* (Imai *et al.,* 2000). The present analysis identified 108 independent genes that were expressed specifically in endodermal cells (Table 3). Many endoderm-specific genes in the larva belong to subclass AV, which includes protein synthesis cofactors, tRNA synthetase, and ribosomal proteins (Table 4; Fig. 1). Although we categorized these genes as "endoderm-specific," we cannot exclude the possibility that they are also expressed in some mesodermal cells in the trunk. One typical example is cluster 00320, which encodes a protein similar to ribosomal protein S3a. This gene is apparently expressed in all cells of the trunk except for those of the epidermis and nervous system (Fig. 3A). In the ascidian larva, protein synthesis is probably most active in endodermal and mesodermal cells that are precursors of adult tissues and organs (cf., Hirano and Nishida, 1997, 2000).

Besides genes expressed in the whole trunk endoderm as described above, several endoderm-specific genes show intriguing expression patterns. Cluster 00016, encoding a protein similar to *Locusta migratoria* apolipophorin precursor, was exclusively expressed in the trunk endoderm and endodermal strand (Figs. 3B and 3B′). On the other hand, cluster 00532, showing no significant sequence similarity to known genes, was specifically expressed in the trunk endoderm and CNS, including the brain vesicle, visceral ganglion, and nerve cord, but was not expressed in the endodermal strand (Figs. 3C and 3C′). Several genes are differentially expressed within the trunk endoderm. For example, cluster 00944, which shows no significant matches in the databases, was expressed only in the anterior half of the trunk endoderm (Fig. 3D). Similarly, cluster 00194, encoding a protein similar to ribosomal protein S18, showed a gradient of expression along the anteroposterior axis in the trunk endoderm: stronger in the anterior endoderm and weaker in the posterior endoderm (Figs. 3E and 3E′). This gene was also expressed in the endodermal strand. Two genes, 00859 (Fig. 3F) and 02129 (data not shown), both with no significant sequence similarity to known genes, were predominantly expressed in the middle part of the trunk endoderm, and the expression of these genes was weaker or absent in both the anterior and posterior parts of the endoderm.

A particularly intriguing gene expression pattern was shown by cluster 00286, which encodes a protein similar to mammalian SOUL protein (Figs. 3G and 3H). This gene was expressed in the posterior endoderm and endodermal strand. The cells expressing cluster 00286 were limited to only a portion of the endodermal strand, and the expression region varied among individual larvae with respect to the anteroposterior axis (compare Fig. 3G with Fig. 3H). Although these larvae had been fertilized at the same time and ascidian embryos develop synchronously, progression of development may become less synchronized in larval stages

as seen in variation of timing of hatching and onset of metamorphosis. Therefore, the variation observed in the cluster 00286 may represent dynamically changing gene expression patterns during larval stages. A less plausible alternative is that cells expressing gene 00286 are migrating along the endodermal strand in the larva.

The trunk endoderm of the larva develops into a variety of endodermal organs in the adult, including the peribranchial epithelium, branchial sac, endostyle, esophagus, stomach, and intestine (Hirano and Nishida, 2000). The heterogeneity of endoderm-specific gene expression observed in the present study may correspond to the developmental fates of endodermal cells. A detailed cell-lineage analysis of adult endodermal organs suggested that cell-fate specification within the endoderm is a position-dependent rather than a deterministic and lineage-based process (Hirano and Nishida, 2000). Endoderm-specific genes found in this study should be useful tools for investigating developmental mechanisms of endodermal organs in adult ascidians.

MESENCHYME. The *Ciona* tadpole larva contains four pockets (two pairs) of mesodermal cells (mesenchyme) situated in the posterior part of the trunk (e.g., Katz, 1983). Recent analysis of gene expression profiles in *Ciona* tailbud embryos identified 111 genes that were expressed exclusively in mesenchyme cells (Satou *et al.*, 2001). In the present study, 34 out of 1013 clusters were expressed exclusively in mesenchyme cells (Table 3).

Interestingly, mesenchyme-specific genes are usually expressed not only in the posterior part of the larval trunk, but also in a small area of the anterior trunk near the adhesive organ. For example, genes 00141 (Fig. 3I) and 00389 (Fig. 3J), encoding proteins similar to oxidoreductase and membrane glycoprotein 25L, respectively, are expressed in posterior mesenchyme cells as well as a group of cells in the anterior part of the trunk. In tailbud embryos, these genes are specifically expressed in mesenchyme cells. One plausible explanation for this expression pattern is that mesenchyme cells migrate anteriorly during larval development. Alternatively, anterior cells may acquire *de novo* the expression of these genes. Because the mesenchyme of the ascidian larva is a source of adult mesodermal tissues, it will be interesting to follow the expression of mesenchyme-specific genes identified in the present study during metamorphosis in order to understand the development of adult mesodermal tissues.

NOTOCHORD CELLS. The notochord of an ascidian larva is composed of just 40 cells, and the entire lineage has been completely described (Nishida, 1987). Previous studies revealed that ascidian *Brachyury* genes, *HrBra* (*As-T*) in *H. roretzi* (Yasuo and Satoh, 1993) and *Ci-Bra* in *C. intestinalis* (Corbo *et al.*, 1997), are expressed exclusively in the notochord precursor cells and play a pivotal role in notochord formation (Yasuo and Satoh, 1998). Isolation and characterization of *Ci-Bra*-downstream genes demonstrated that nearly 40 genes are expressed specifically and predominantly in notochord cells (Takahashi *et al.*, 1999; Hotta *et al.*, 2000).

The present analysis revealed that 23 genes were specifically expressed in the notochord (Table 3). Although some of these genes have previously been identified as notochord-specific genes, others were newly identified in the present analysis. For example, cluster 00185 (Figs. 3K and 3K′), encoding a protein similar to osteonectin, was expressed exclusively in the notochord, as reported in tailbud embryos (Satou *et al.*, 2001), while the expression pattern of cluster 02103 (Fig. 3L), with significant similarity to caveolin, was not examined in previous analyses. In the present analysis, 18 genes were newly identified as notochord-specific genes. To fully elucidate the genetic program governing notochord development, it should be examined whether or not these genes are targets of the *Ci-Bra* transcription factor.

MUSCLE CELLS. The *Ciona* tailbud embryo develops 18 unicellular, striated muscle cells on each side of the tail (total 36 cells); of them, 28 cells in the anterior and middle part of the tail are derived from the B4.1 pair (primary lineage), while 4 cells in the posterior part and 4 cells at the tip of the tail originate from the A4.1 and b4.2 pairs, respectively (Nishida, 1987). Several muscle-specific genes, including genes for actin (Kusakabe *et al.*, 1991, 1995) and myosin heavy chain (Makabe and Satoh, 1989; Araki and Satoh, 1996), have been characterized in *Halocynthia* embryos. In addition, cDNA clones for troponin I (MacLean *et al.*, 1997) and tropomyosin (Meedel and Hastings, 1993) have been isolated from *C. intestinalis*.

The present study revealed that 28 genes were specifically expressed only in muscle cells. As expected, numerous (12 of 28) muscle-specific genes encode proteins for muscle contraction. For example, cluster 00327 encodes myosin heavy chain (Fig. 3M), cluster 00359 encodes troponin T, cluster 01464 encodes myosin-binding protein C, cluster 04951 encodes a ryanodine receptor, and cluster 05749 encodes troponin C. In addition to genes expressed exclusively in muscle cells, several genes were specifically expressed both in the muscle and the nervous system of the larva. For example, cluster 01974, encoding creatine kinase-M, was conspicuously expressed both in muscle cells and the CNS (Figs. 3N and 3N′).

All of the muscle-specific genes examined in this study were expressed in every muscle cell, with the exception of one gene: cluster 03396, which shows no significant sequence similarity to known genes and was expressed only in muscle cells of the posterior part of the tail (Fig. 3O). In the previous analysis of gene expression profiles in *Ciona* tailbud embryos, one gene, 01633, with no significant matches in the databases, showed the same expression pattern (Satou *et al.*, 2001). It will be interesting to determine molecular mechanisms underlying the spatial expression of these two genes in future studies.

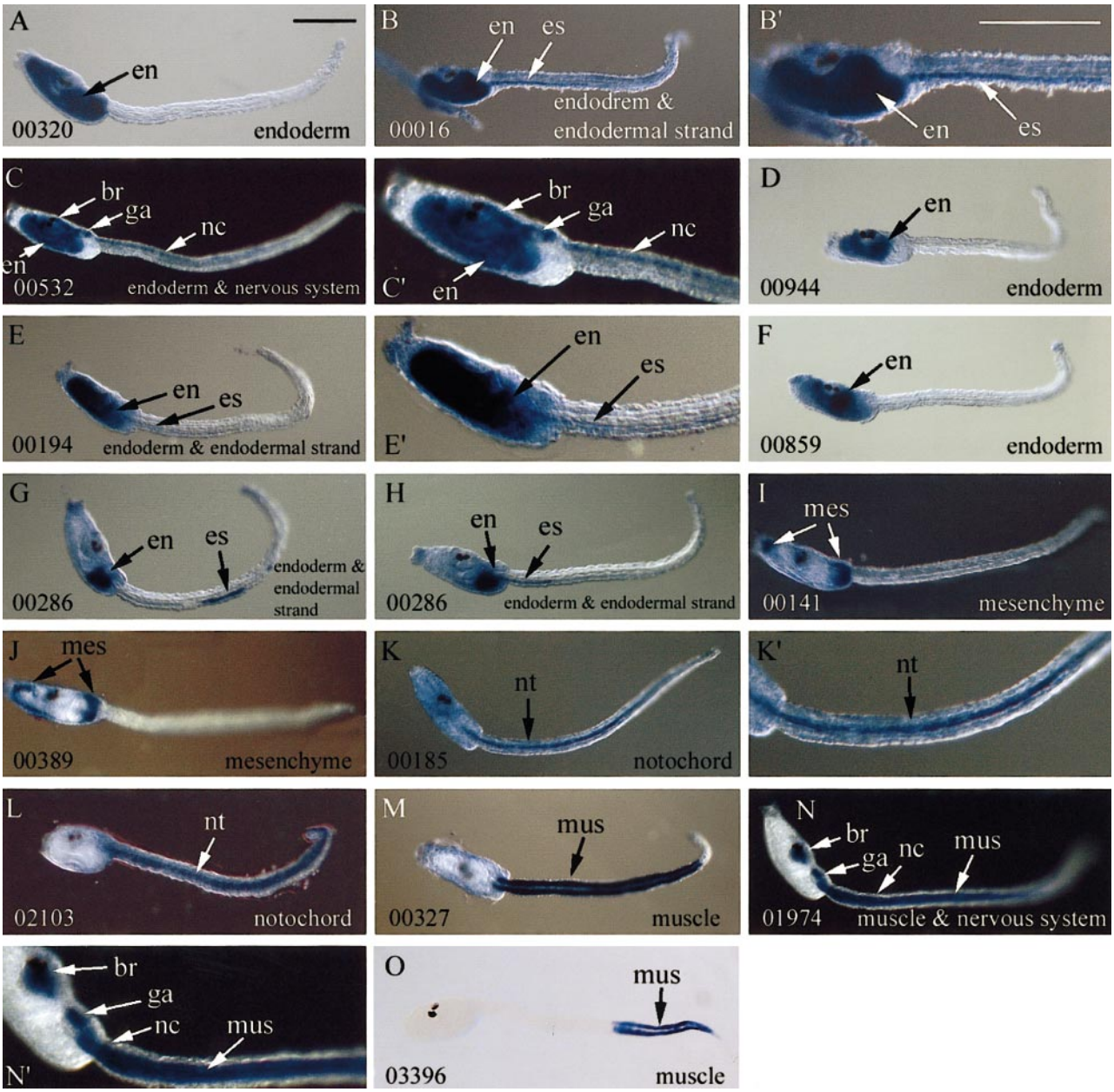

FIG. 3. Whole-mount *in situ* hybridization showing the expression profiles of genes specific to organs and/or tissues in *C. intestinalis* larvae. The organ where the gene is specifically expressed is shown in the right bottom corner, and the clone names in the left bottom corner. (B′), (C′), (E′), (K′), and (N′) are enlargements of (B), (C), (E), (K), and (N), respectively. br, brain; en, endoderm; es, endodermal strand; ga, visceral ganglion; mes, mesenchyme; mus, muscle; nc, nerve cord; nt, notochord. Scale bars in (A) and (B′) are 100 μm and are applicable to all photographs. See text for details.

RUDIMENTS OF ADULT ORGANS AND OTHER UNIDENTIFIED CELLS. The adult ascidian has two siphons, the oral siphon and the atrial siphon. The oral siphon rudiment is located in the dorsal epidermis anterior to the brain vesicle, while the two atrial primordia appear as pockets of cuboidal and columnar cells in the dorsal epidermis of the posterior part of the trunk (Katz, 1983). During metamorphosis, the two atrial primordia fuse to form a single atrial siphon (Berrill, 1950). A recent study identified a gene, *Ci-meta2*, that is expressed in the siphon rudiments of *Ciona* larvae (Nakayama *et al.*, 2001). The *Ci-meta2* gene encodes a protein with a putative secretion signal and three thrombospondin type-1 repeats.

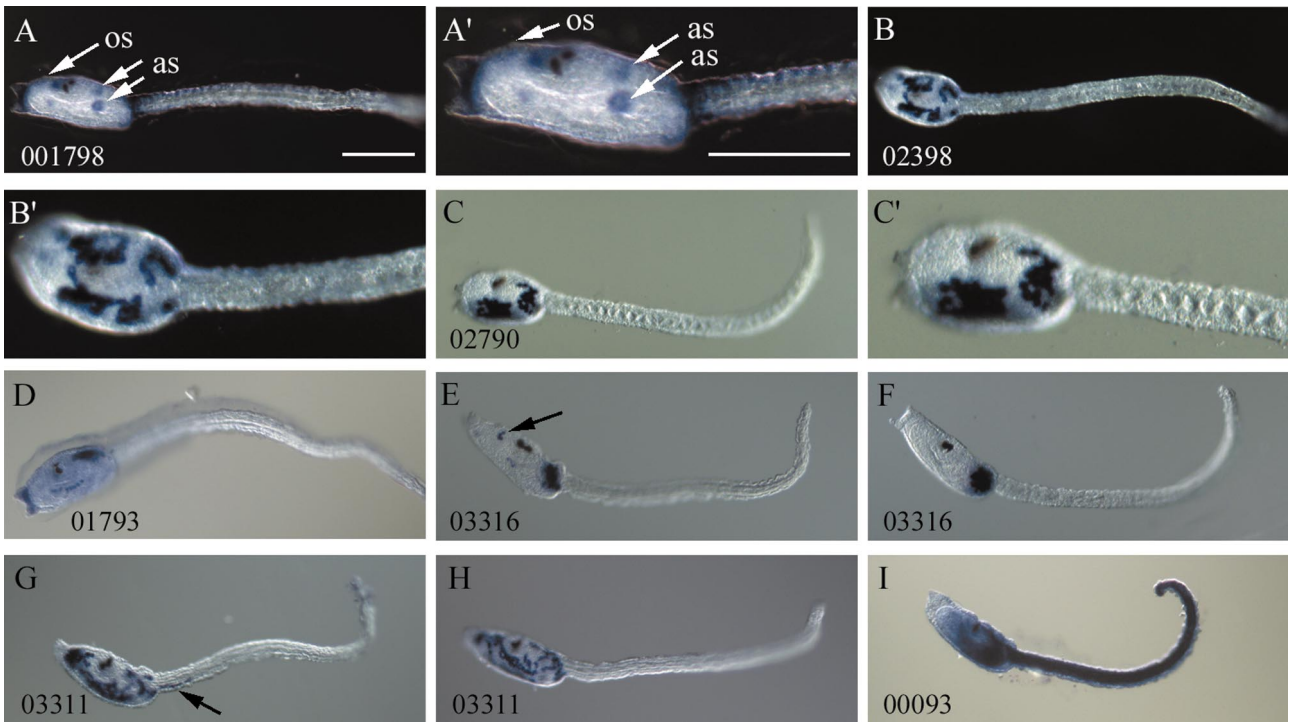

FIG. 4. Whole-mount *in situ* hybridization showing the expression profiles of genes specific to organs and/or tissues in *C. intestinalis* larvae. The clone names are shown in the left bottom corner. (A′), (B′), and (C′) are enlargements of (A), (B), and (C), respectively. as, atrial siphon; os, oral siphon. Scale bars in (A) and (A′) are 100 μm and are applicable to all photographs. See text for details.

Ci-meta2 expression begins at the larval stage in the adhesive organ, the neck region between the trunk and the tail of the larva, and the siphon primordia, and the expression is upregulated in the metamorphosing juveniles.

In the present analysis, eight genes showed an expression pattern similar to that of *Ci-meta2*. For example, cluster 01798, encoding a protein similar to diacyl glycerol kinase, was expressed in the oral and atrial siphon rudiments, the adhesive organ, and the epidermis of the neck region between the trunk and tail (Figs. 4A and 4A′). We also identified genes expressed only in the oral siphon primordium (cluster 04783) or the atrial siphon primordia (clusters 04765 and 04798). The expression patterns of these genes suggest that they play a role in morphogenesis and remodeling of epidermal cells during metamorphosis.

The present *in situ* hybridization analysis revealed that a number of genes, most of which display no significant sequence similarity to known genes, were expressed in unidentified cells scattered in the trunk of the larva. For example, clusters 02398 (Figs. 4B and 4B′) and 02790 (Figs. 4C and 4C′) were expressed in bilaterally distributed cell populations in the ventral region of the trunk. Cluster 01793 was expressed in a group of cells in the posterior–dorsal region of the trunk and in bilateral rows of cells in the endodermal region of the trunk (Fig. 4D). The posterior cells expressing this gene seem to correspond to the position of the trunk lateral cells. In *Halocynthia* larvae, the trunk lateral cells migrate anteriorly and are scattered in the entire trunk region (Nishide *et al.*, 1989). Therefore, cells expressing gene 01793 are probably trunk lateral cells of *Ciona* larvae. A similar expression pattern was observed for cluster 03316 (Figs. 4E and 4F). There was some variation in the expression pattern of this gene: in some embryos, a small number of cells expressing 03316 were distributed in trunk regions anterior or ventral to the brain vesicle (Fig. 4E), but the other embryos showed no such expression (Fig. 4F). The expression pattern of gene 03316 suggests migration of cells from the posterior trunk to the anterior regions during larval development, and different expression patterns of this gene may be dependent on different developmental stages during the period of about 12 h constituting the swimming larva stage.

Variation of the gene expression pattern was also observed for gene 03311. Cells expressing gene 03311 were scattered under the epidermis of the whole trunk (Figs. 4G and 4H). In addition, some embryos showed the expression in the anterior-most part of the endodermal strand of the tail (Fig. 4G, arrow). Although their cell lineages and developmental fates have not been precisely determined yet, cells showing hybridization signals in Figs. 4B–4H are likely to be precursors of adult tissues, belonging to endodermal, mesenchymal, trunk–lateral, or trunk–ventral

cell lineages. Therefore, future studies on the expression and function of these genes will contribute to understanding the development of mesodermal and endodermal tissues in ascidian adults.

The present analysis also identified genes expressed in many kinds of tissues in a position-specific (rather than tissue-specific) manner. For example, cluster 00093, encoding a collagen α1 chain, was expressed in almost all tissue types, but the expression was very weak or absent in the anterior part of the trunk (Fig. 4I). This gene showed basically the same expression pattern in tailbud embryos (Satou *et al.*, 2001). It will be interesting to investigate the molecular and cellular bases of this type of gene expression in terms of the positional identity of cells in the tadpole larva.

In conclusion, analysis of a set of 1013 independent clusters of cDNA clones derived from *Ciona intestinalis* larvae revealed that 545 of them showed significant matches to reported proteins, while 153 did not have enough information to be categorized and 315 did not show significant similarities to any known proteins. Sequence similarity analyses of the 545 clusters suggest that 407 of them are associated with functions that many kinds of cells use, 104 with cell–cell communication, and 34 with transcription factor functions or other gene regulatory proteins. Whole-mount *in situ* hybridization analysis of all of the 1013 clusters demonstrated that a total of 361 clusters showed gene expression that was specific to a single tissue or organ. In addition, 190 genes were specifically expressed in multiple tissues. The present analysis also highlighted characteristic gene expression profiles that were dependent on the type of larval tissues. These data therefore provide new information about many molecular markers for every tissue and organ in the *Ciona* larva, including functional larval organ systems (epidermis, notochord, tail muscle, and nervous system) and adult organ rudiments (mesenchyme, trunk lateral cells, trunk ventral cells, endoderm, and siphon primordia). The nucleotide sequence information obtained here will also be used for future whole-genome analysis to explore molecular mechanisms involved in the development of one of the most primitive chordate body plans.

ACKNOWLEDGMENTS

We thank Hisayoshi Ishikawa, Chikako Imaizumi, Kazuko Hirayama, and Reiko Nakamura for technical support. This research was supported by a Grant-in-Aid for Priority Area C (No. 12202001) from the Ministry of Education, Science, Sports, Culture and Technology, Japan (to N.S.). The Academia DNA Sequencing Center is supported by a Grant-in-Aid for Scientific Research on Priority Area C (No. 12201001) from the same Ministry (to Y.K.).

REFERENCES

Araki, I., and Satoh, N. (1996). *cis*-Regulatory elements conserved in the proximal promoter region of an ascidian embryonic muscle myosin heavy-chain gene. *Dev. Genes Evol.* **206,** 54–63.

Berrill, N. J. (1950). "The Tunicata with an Account of the British Species." Ray Society, London.

Chiba, S., Satou, Y., Nishikata, T., and Satoh, N. (1998). Isolation and characterization of cDNA clones for epidermis-specific and muscle-specific genes in *Ciona savignyi* embryos. *Zool. Sci.* **15,** 239–246.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162,** 156–159.

Corbo, J. C., Levine, M., and Zeller, R. W. (1997). Characterization of a notochord-specific enhancer from the *Brachyury* promoter region of the ascidian *Ciona intestinalis*. *Development* **124,** 589–602.

Crowther, R. J., and Whittaker, J. R. (1992). Structure of the caudal neural tube in an ascidian larva: Vestiges of its possible evolutionary origin from a ciliated band. *J. Neurobiol.* **23,** 280–292.

Di Gregorio, A., and Levine, M. (1998). Ascidian embryogenesis and the origins of the chordate body plan. *Curr. Opin. Genet. Dev.* **8,** 457–463.

Fujiwara, S., Maeda, Y., Shin-i, T., Kohara, Y., Takatori, N., Satou, Y., and Satoh, N. (2002). Gene expression profiles in *Ciona intestinalis* cleavage-stage embryos. *Mech. Dev.*, in press.

Hirano, T., and Nishida, H. (1997). Developmental fates of larval tissues after metamorphosis in ascidian *Halocynthia roretzi*. I. Origin of mesodermal tissues of the juvenile. *Dev. Biol.* **192,** 199–210.

Hirano, T., and Nishida, H. (2000). Developmental fates of larval tissues after metamorphosis in the ascidian *Halocynthia roretzi*. II. Origin of endodermal tissues of the juvenile. *Dev. Genes Evol.* **210,** 55–63.

Hotta, K., Takahashi, H., Asakura, T., Saitoh, B., Takatori, N., Satou, Y., and Satoh, N. (2000). Characterization of *Brachyury*-downstream notochord genes in the *Ciona intestinalis* embryo. *Dev. Biol.* **224,** 69–80.

Hudson, C., and Lemaire, P. (2001). Induction of anterior neural fates in the ascidian *Ciona intestinalis*. *Mech. Dev.* **100,** 189–203.

Imai, K., Takada, N., Satoh, N., and Satou, Y. (2000). β-Catenin mediates the specification of endoderm cells in ascidian embryos. *Development* **127,** 3009–3020.

Ishida, K., Ueki, T., and Satoh, N. (1996). Spatio-temporal expression patterns of eight epidermis-specific genes in the ascidian embryo. *Zool. Sci.* **13,** 699–709.

Katsuyama, Y., Sato, Y., Wada, S., and Saiga, H. (1999). Ascidian tail formation requires *caudal* function. *Dev. Biol.* **213,** 257–268.

Katz, M. J. (1983). Comparative anatomy of the tunicate tadpole *Ciona intestinalis*. *Biol. Bull.* **164,** 1–27.

Kumano, G., and Nishida, H. (1998). Maternal and zygotic expression of the endoderm-specific alkaline phosphatase gene in embryos of the ascidian *Halocynthia roretzi*. *Dev. Biol.* **198,** 245–252.

Kusakabe, T., Suzuki, J., Saiga, H., Jeffery, W. R., Makabe, K. W., and Satoh, N. (1991). Temporal and spatial expression of a muscle actin gene during embryogenesis of the ascidian *Halocynthia roretzi*. *Dev. Growth Differ.* **33,** 227–234.

Kusakabe, T., Hikosaka, A., and Satoh, N. (1995). Coexpression and promoter function in two muscle actin gene complexes of different structural organization in the ascidian *Halocynthia roretzi*. *Dev. Biol.* **169,** 461–472.

Lee, Y.-H., Huang, G. M., Cameron, R. A., Graham, G., Davidson, E. H., Hood, L., and Britten, R. J. (1999). EST analysis of gene

expression in early cleavage-stage sea urchin embryos. *Development* **126,** 3857–3867.

MacLean, D. W., Meedel, T. H., and Hastings, K. E. (1997). Tissue-specific alternative splicing of ascidian troponin I isoforms. Redesign of a protein isoform-generating mechanism during chordate evolution. *J. Biol. Chem.* **272,** 32115–32120.

Makabe, K. W., and Satoh, N. (1989). Temporal expression of myosin heavy chain gene during ascidian embryogenesis. *Dev. Growth Differ.* **31,** 71–77.

Meedel, T. H., and Hastings, K. E. M. (1993). Striated muscle-type tropomyosin in a chordate smooth muscle, ascidian body-wall muscle. *J. Biol. Chem.* **268,** 6755–6764.

Nakayama, A., Satou, Y., and Satoh, N. (2001). Isolation and characterization of genes that are expressed during *Ciona intestinalis* metamorphosis. *Dev. Genes Evol.* **211,** 184–189.

Nicol, D., and Meinertzhagen, I. A. (1991). Cell counts and maps in the larval central nervous system of the ascidian *Ciona intestinalis* (L.). *J. Comp. Neurol.* **309,** 415–429.

Nishida, H. (1987). Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. III. Up to the tissue restricted stage. *Dev. Biol.* **121,** 526–541.

Nishida, H. (1991). Induction of brain and sensory pigment cells in the ascidian embryo analyzed by experiments with isolated blastomeres. *Development* **112,** 389–395.

Nishide, K., Nishikata, T., and Satoh, N. (1989). A monoclonal antibody specific to embryonic trunk-lateral cells of the ascidian *Halocynthia roretzi* stains coelomic cells of juvenile and adult basophilic blood cells. *Dev. Growth Differ.* **31,** 595–600.

Nishikata, T., Yamada, L., Mochizuki, Y., Satou, Y., Shin-i, T., Kohara, Y., and Satoh, N. (2001). Profiles of maternally expressed genes in fertilized eggs of *Ciona intestinalis*. *Dev. Biol.* **238,** 315–331.

Nishino, A., and Satoh, N. (2001). The simple tail of chordates: Phylogenetic significance of appendicularians. *Genesis* **29,** 36–45.

Okado, H., and Takahashi, K. (1988). A simple "neural induction" model with two interacting cleavage-arrested ascidian blastomeres. *Proc. Natl. Acad. Sci. USA* **85,** 6197–6201.

Pearson, W. R., and Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85,** 2444–2448.

Reverberi, G., Ortolani, G., and Farinella-Ferruzza, N. (1960). The causal formation of the brain in the ascidian larva. *Acta Embryol. Morphol. Exp.* **3,** 296–336.

Satoh, N. (1994). "Developmental Biology of Ascidians." Cambridge Univ. Press, New York.

Satoh, N. (1999). Cell fate determination in the ascidian embryo. In "Cell Lineage and Fate Determination" (S. A. Moody, Ed.), pp. 59–74. Academic Press, New York.

Satoh, N. (2001). Ascidian embryos as a model system to analyze expression and function of developmental genes. *Differentiation* **68,** 1–12.

Satoh, N., and Jeffery, W. R. (1995). Chasing tails in ascidians: Developmental insights into the origin and evolution of chordates. *Trends Genet.* **11,** 354–359.

Satou, Y., Takatori, N., Yamada, L., Mochizuki, Y., *et al.* (2001). Gene expression profiles in *Ciona intestinalis* tailbud embryos. *Development* **128,** 2893–2904.

Satou, Y., Takatori, N., Fujiwara, S., Nishikata, T., *et al.* (2002). *Ciona intestinalis* cDNA projects: EST analyses and gene expression profiles during embryogenesis. *Gene*, in press.

Simmen, M. W., Leitgeb, S., Clark, V. H., Jones, S. J. M., and Bird, A. (1998). Gene number in an invertebrate chordate *Ciona intestinalis*. *Proc. Natl. Acad. Sci. USA* **95,** 4437–4440.

Tabara, H., Motohashi, T., and Kohara, Y. (1996). A multi-well version of in situ hybridization on whole mount embryos of *C. elegans*. *Nucleic Acids Res.* **24,** 2119–2124.

Takahashi, H., Hotta, K., Erives, A., Di Gregorio, A., Zeller, R. W., Levine, M., and Satoh, N. (1999). *Brachyury* downstream notochord differentiation in the ascidian embryo. *Genes Dev.* **13,** 1519–1523.

Takamura, K. (1998). Nervous network in larvae of the ascidian *Ciona intestinalis*. *Dev. Genes Evol.* **208,** 1–8.

Ueki, T., Makabe, K. W., and Satoh, N. (1991). Isolation of cDNA clones for epidermis-specific genes of the ascidian embryo. *Dev. Growth Differ.* **33,** 579–586.

Ueki, T., and Satoh, N. (1995). Sequence motifs shared by the 5′ flanking regions of two epidermis-specific genes in the ascidian embryo. *Dev. Growth Differ.* **37,** 597–604.

Whittaker, J. R. (1990). Determination of alkaline phosphatase expression in endodermal cell lineages of an ascidian embryo. *Biol. Bull.* **178,** 222–230.

Yasuo, H., and Satoh, N. (1993). Function of vertebrate *T* gene. *Nature* **364,** 582–583.

Yasuo, H., and Satoh, N. (1998). Conservation of the developmental role of *Brachyury* in notochord formation in a urochordate, the ascidian *Halocynthia roretzi*. *Dev. Biol.* **200,** 158–170.

Received for publication September 4, 2001
Revised October 22, 2001
Accepted November 7, 2001
Published online January 15, 2002